MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2017/2018

TIB2341 – INTRODUCTION TO BIOINFORMATICS

(All sections / Groups)

30 May 2018
2:30-4:30 PM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use the calculator.

QUESTION 1 [10 MARKS]

a) What is gene expression? [1 mark]

b) Based on the nucleotide base pairing concept, state the mRNA strand for the template DNA strand given. Please indicate the orientation of mRNA.
Template DNA strand: 5'-CCTAGCTA-3' [2 marks]

c) Describe briefly the concept of genetic code redundancy. [2 marks]

d) Name any TWO shape of secondary structure of protein. [1 mark]

e) What is the purpose of sequence alignment? [1 mark]

f) What is FASTA format? [1 mark]

g) What are primary and secondary biological databases? Give ONE example for each. [2 marks]

QUESTION 2 [10 MARKS]

a) Name ONE protein sequence database and ONE protein structure database commonly used for biological research. [1 mark]

b) Name TWO bioinformatics applications for the biomolecule sequence analysis. [1 mark]

c) If you want to search for biological research literature information, what would be a good website to visit? [1 mark]

d) What are the criteria for deciding if a sequence is a CpG island? Why such CpG sequences are called "islands"? Why are they biologically interesting? [2 marks]

e) What are the four protein structure classifications? Explain. [4 marks]

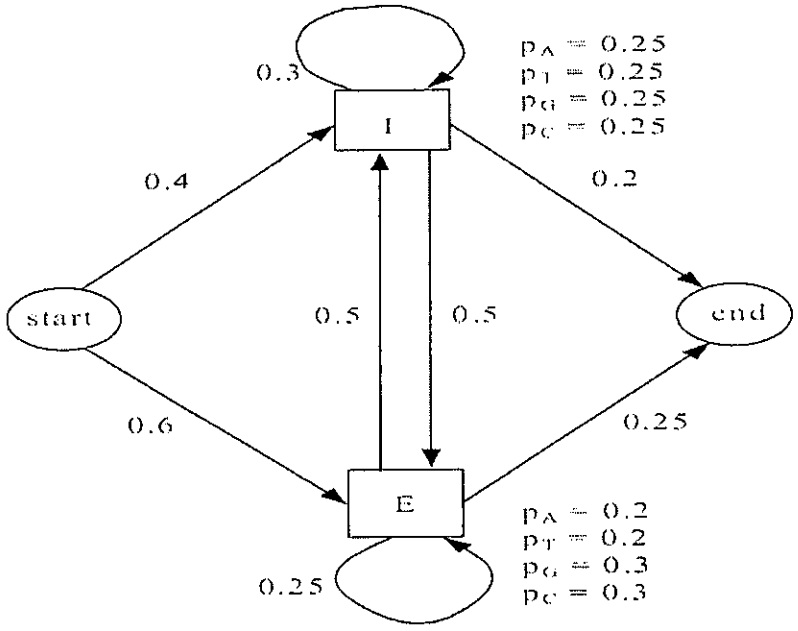
f) Calculate the possible number of rooted and unrooted trees with eight taxa. [1 mark]

Continued...

QUESTION 3 [10 MARKS]

a) Using the dynamic programming method, construct the local alignment score table for the following two sequences and find the possible alignment. The scoring parameters are:
Match score = +1, mismatch score = -1, gap penalty = -2 [5 marks]
CTGTAGG
GTGTAC

b) Find the probability of the DNA segment "TCA" predicted by the Hidden Markov Model with the state-transition diagram given in the figure below, using the **Forward Algorithm** [5 marks]



QUESTION 4 [10 MARKS]

a) Explain the steps in supervised learning to train the ANN for 2 different input signals. [2 marks]

b) List at least four applications of Artificial Neural Network (ANN) in Bioinformatics. [2 marks]

c) What are the two different strategies in reference-based protein structure prediction? [2 marks]

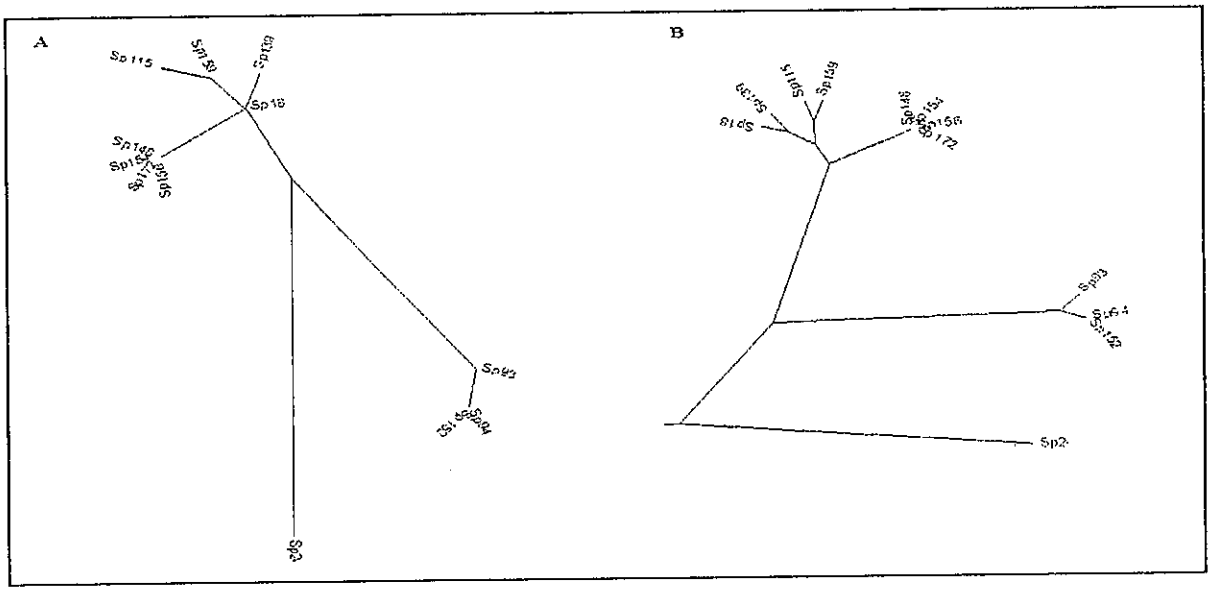
Continued...

d) Define ANN - Artificial Neural Networks. Draw a simple diagram of a 2-layer ANN (input layer, hidden layer, and output). [3 marks]

e) What is one main disadvantage of ab initio protein structure prediction methods? [1 mark]

QUESTION 5 [10 MARKS]

a) The phylogenetic trees below were reconstructed based on agglomerative distance methods. State the possible method used to generate each tree and justify your answer. [4 marks]



b) Given the following distance matrix, construct the corresponding phylogenetic tree using the distance method (UPGMA). [5 marks]

	A	B	C	D	E	F
A	0					
B	5	0				
C	4	7	0			
D	7	10	7	0		
E	6	9	6	5	0	
F	8	11	8	9	8	0

c) How is outlier/outgroup represented in a phylogenetic tree context? [1 mark]

END OF PAPER